# The MY NASA DATA Project: Tools for Knowledge Sharing and Discovery

Lin H. Chambers[1], Erica J. Alston[1], Dennis D. Diones[2], Susan W. Moore[2], Penny C. Oots[2], Carrie S. Phelps[2]
[1]NASA Langley Research Center, MS 420, Hampton, VA 23681-2199
[2]Science Applications International Corporation, Hampton, VA 23666
mynasadata@larc.nasa.gov

**Abstract - The Mentoring and inquirY using NASA Data on Atmospheric and Earth science for Teachers and Amateurs (MY NASA DATA) project (http://mynasadata.larc.nasa.gov) began in 2004, as part of NASA's Research, Education, and Applications Solutions Network (REASoN), in order to open NASA Earth Science data – typically held in large files with specialized file formats - to a broader community including K-12 education and citizen scientists. MY NASA DATA consists of a Web site that collects tools, lesson plans, science project ideas, and specially developed documentation to help the target audiences more easily use NASA's vast collection of data about the Earth System. The core piece of the MY NASA DATA project is the creation of microsets (small datasets both static and custom) that make data easily accessible. The installation of a Live Access Server (LAS) has greatly enhanced the ability to create and explore custom microsets of Earth System Science data by teachers and students, particularly those held at the Atmospheric Science Data Center (ASDC), where the project is based. As an open source software tool using emerging data standards through the Open source Project for a Networked Data Access Protocol (OPeNDAP), LAS also allows the MY NASA DATA team to make available data on other aspects of the Earth System from collaborating data centers and projects.**

## I. INTRODUCTION

The MY NASA DATA project was proposed to the REASoN Cooperative Agreement Notice primarily under the category of Education. However, the proposal included an open source component for the development of tools for data analysis, visualization, and exploration. Thus, the project also receives some attention through the Earth Science Technology Office. In its implementation, the project has concentrated on the use of a single set of open source software, called a Live Access Server. This tool, already well developed, was identified and installed by Aaron Dalton, a Co-operative Education student from Virginia Polytechnic Institute and State University, who spent a summer working with the MY NASA DATA team in 2004. The LAS was found to be very nearly ideal for the goals of the project. Thus, this paper will concentrate on a description of the Live Access Server, and some of the specifics regarding our use of it.

## II. The Live Access Server

The Live Access Server is an open source software tool initially developed under the auspices of the National Oceanic and Atmospheric Administration (NOAA), through its Pacific Marine Environmental Laboratory (PMEL). Current information on the LAS is available on the Internet (http://ferret.wrc.noaa.gov/Ferret/LAS/). The LAS is a "highly configurable web server designed to provide flexible access to geo-referenced scientific data". The LAS provides a number of useful services:

- Custom data visualization, subsetting and reformatting on-the-fly, using a simple graphical user interface (GUI)
- Comparison of two different variables within the GUI
- Unified access to multiple types of data in a single interface
- Networking capabilities with OPeNDAP data servers

The initial installation and testing of our LAS server was completed in just two weeks. Currently, the MY NASA DATA Live Access Server is configured to hold over 100 scientific parameters from six remote sensing projects hosted by ASDC at NASA Langley Research Center in Hampton, Virginia. It can also serve data from remote data sources hosting OPeNDAP servers with suitable data.

The LAS user interface (UI) and backend application are written using a combination of Java modules, Perl scripts and Velocity macros. The setup also requires an Apache web server, Tomcat and a MySQL database. The user interface manages the web browser interface and provides the user the ability to select a dataset and variable to visualize, as well as allowing selection of geographical regions, visualization style and temporal restrictions. Figure 1 shows the configuration of the LAS user interface available on the MY NASA DATA website. The user interface communicates with the backend application using Extensible Markup Language (XML). XML configuration files allow the project to store structured

metadata for the server in a text file similar to HTML, easily manipulated and readable by our web developers.

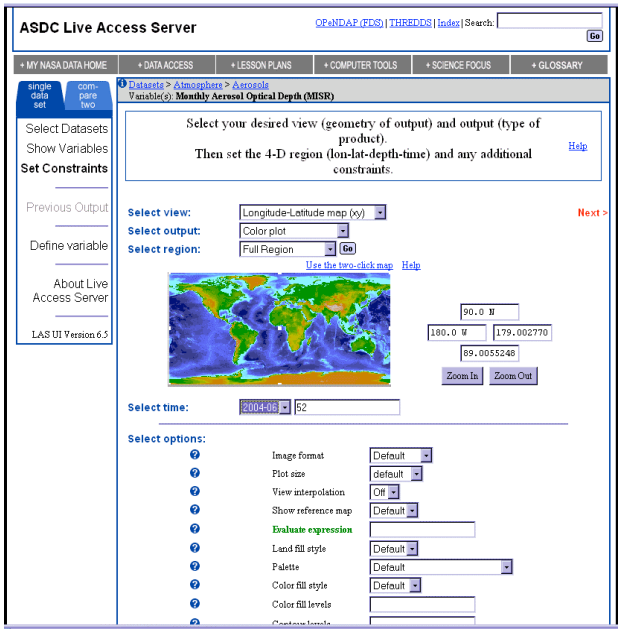


Figure 1. LAS User Interface

The UI calls the backend application to perform data analysis and visualization. The LAS installation package is configured to use Ferret as the data analysis and visualization program, but other programs such as IDL can be used. The Ferret application dynamically creates custom Ferret scripts that are then used by Ferret to visualize or analyze the requested data. This use of Ferret is seamless to the LAS user. The LAS uses popup windows to display the results of the analysis and visualization application processing. The visualized or analyzed data can be returned to the user in a variety of formats such as a GIF image, ArcView gridded, comma separated values, or plain ASCII text. An example of an output image is given in Figure 2. LAS users should be made aware that popup blockers could render users unable to see results. Newer versions of LAS allow browsers with medium level pop up to view the LAS output window.

The LAS is designed to serve gridded data where data points are fairly regularly distributed in time and space. This type of data product is representative of many of the holdings selected from ASDC data for the MY NASA DATA website. Datasets can be hosted on the local machine or can be accessed remotely since LAS is also an OPeNDAP client. As a client, LAS can send URL requests for specific data through the network to an OPeNDAP server that answers with the requested data. OPeNDAP can allow access to data that is formatted in many different ways. For example, OPeNDAP can access netCDF, HDF, and JGOFS formatted data. The user need not learn about any of the formats, since the OPeNDAP server and client operate to deliver the requested data in the format in which the analysis package expects to see it. Using OPeNDAP, only the requested suset is returned, and unnecessary data are not sent over the Internet. However, when using OPeNDAP, the remote site must be monitored to ensure it is operating. For instance, during Hurricane Katrina, the Naval Oceanographic Office (NAVOCEANO) server that was providing sea surface temperature data via OPeNDAP unfortunately went down and the dataset had to be temporarily removed from the LAS.

The Live Access Server is most basically implemented if using data in netCDF (network Common Data Form) format. This format is a set of interfaces for array-oriented data access and was the easiest for us to implement initially. LAS supports a methodology for presenting multiple netCDF files as a single continuous time series and so has the effect of treating multiple netCDF files as one virtual file. The MY NASA DATA project team also wrote code to translate scientific data from HDF format to netCDF format prior to the installation and testing of our OPeNDAP HDF server. Ferret can also be used in standalone mode to produce netCDF files from ASCII text or binary files. This provides a straight-forward way to prepare these type of data to be served by LAS. Once properly configured, gridded datasets can easily be served, but we have also encountered some datasets that do not contain some identifying information within the dataset. For instance, if the time coverage of the file is designated by the file name only, some processing is necessary to define the time period available to the LAS.

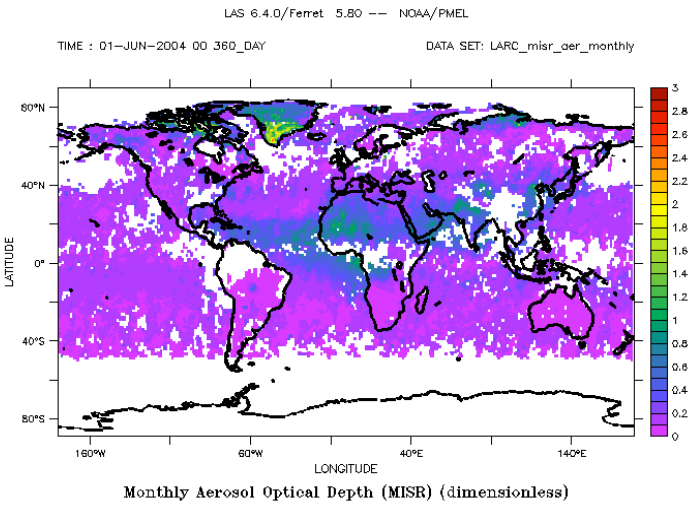


Figure 2. LAS sample output

## III. LAS and MY NASA DATA

The MY NASA DATA team is using the LAS to make available selected parameters from scientific data sets for use by our target audiences: the K-12 education community, and the citizen science community (we define the latter broadly to include K-12 students doing science projects). The

parameters are carefully selected from larger scientific datasets to be things that our target audience can readily understand. We have organized the data within the LAS following the data structure of the Global Change Master Directory (GCMD; http://gcmd.nasa.gov/). To date we have 127 parameters available under Atmosphere (the primary focus of the Langley Atmospheric Sciences Data Center), Cryosphere (Snow and Ice), and Surface. Figure 3 shows a list of currently available topics and parameters.


Figure 3. MY NASA DATA LAS parameters

### A. LAS Documentation

One of the primary challenges of implementing an LAS server for the MY NASA DATA target audiences has been the issue of documentation. We have used the existing LAS help features – an icon that appears at various steps through the data selection pages – to insert custom documentation for our target audience. At this time, we mainly focus on defining the complex scientific parameters served through the LAS in our illustrated MY NASA DATA Science Glossary. The standard LAS help pages also remain available, through a Help link.

In addition to this documentation within LAS, experience with a diverse group of educators in a variety of workshop settings over the last two years has demonstrated that additional documentation is required to introduce and explain the LAS. Thus, we have recently developed and posted a customized "Live Access Server Introduction" document (http://mynasadata.larc.nasa.gov/LASintro.html) on the MY NASA DATA site. This document begins with a layman's introduction to the LAS, using the analogy of a customer ordering food in a restaurant. It then provides a brief explanation (menu) of what is available within the LAS, with instructions and links to find a variety of additional information. One special challenge we have is that the datasets within the LAS – mainly obtained from satellite – are not all for the same time period. Thus, we provide a link to a "Time Coverage at a Glance" chart, which provides a potential LAS user the big picture of what they can find and use within LAS.

We also provide a key to the often-confusing parameter names within the LAS. This key is a standard naming convention that we have imposed on the parameters we offer through the LAS, and consistently provides the same type of information in the same order. This naming standard is of considerable help to our audience in navigating through these datasets, and is in many cases a substantial change from the naming convention of the original scientific data set. Currently, we are working on adding information about the data sources (mostly satellite instruments) to this page, to further increase the comfort level of our audience. Finally, this page offers a link to a quick PowerPoint tutorial, which shows screenshots of the LAS GUI, with an explanation of what information needs to be specified at each step. This tutorial has been very well-received by teachers in our workshops.

### B. LAS Data

A continued challenge in implementing the LAS for MY NASA DATA has been the challenge of obtaining and making available sufficient data to enable educators and citizen scientists to explore what they want to explore. The initial core of data within the LAS was obtained from data holdings of the ASDC. We continue to explore new possibilities for data additions from that source. Additional data have been made available for our site through the efforts

of the Global Energy and Water Cycle Experiment (GEWEX) Radiative Flux Assessment (RFA) activity. This international assessment effort is using standard subsets of all available radiation budget parameters, and provides a nice starting point for ingest of selected parameters to our LAS system. These data are publicly available in other forms, but again the challenge of large files and specialized formats makes it easier for us to use the GEWEX RFA subset versions.

In addition, by popular demand (related to El Nino and other topics), we have sought out another source of sea surface temperature (SST) data. We have been in discussion with the Physical Oceanography Data Center at JPL to obtain some ocean datasets. However, we have discovered that they need to make some format changes before we can use their data via OPeNDAP. In the meantime, we are still acquiring SST data through the Naval Oceanographic Office (NAVOCEANO) and currently have more than a month of near real-time data available. Additional data are currently being added through a semi-manual process that involves an ftp push and additional processing on our end since their OPeNDAP server has become unsupported.

### C. *LAS Uses*

While the LAS is open for free exploration, we find that most people need some stepping stones in the learning curve. In addition to the LAS introduction, mentioned above, we provide a number of lesson plans which use specific LAS data, and provide step-by-step instructions for obtaining it. Similarly, a set of project ideas is in development, which will point more generally to relevant LAS parameters.

### D. *LAS Instruction*

The final challenge for implementation of LAS for MY NASA DATA is to make users feel comfortable using this new technology. Two summer teacher workshops, conferences and numerous short courses have enabled the MND team to gain personal interaction and experience with our target audiences. Participants are allowed to explore the LAS and its data at every session, always sparking ideas for lesson plans and eliciting scientific questions. Feedback has been very positive and productive regarding LAS:

> "Sixteen participants reacted favorably to this activity: one stated that it was a good amount of time for the activity, and two said that it was fun and easy to access."

> "My data management class used the data in a self-directed project. They came up with a premise, developed a hypothesis, tested their hypothesis using the microsets then they wrote a conclusion and presented a presentation. In another class we used cloud cover microsets with the lessons I developed at the workshop. I also used the lesson on Hadley cells developed for use without computers. Students came away with a new respect for data analysis, and a new enthusiasm for the subject. All were motivated to use real data, and showed great interest in the activities. I feel that my grade nines came away with a greater understanding of Hadley cells and the relationship between cloud coverage and ground cover. My data management class learned that there is real data everywhere and all must be suspect for validity. I think that the lessons we completed were successful."

For the benefit of the greater WWW audience, our team contact information, an FAQ page, and a link to our e-Mentor Network are also available on the MND website to assist in LAS instruction and scientific data use questions.

## IV. CONCLUSIONS

The MY NASA DATA team has successfully implemented a Live Access Server tool to enable data access and exploration by the K-12 and citizen science communities. This is a powerful tool in our goal to make NASA Earth Science data accessible to these communities, thus enabling knowledge sharing and discovery. Our metrics (Fig. 4)show an ever- increasing amount of data exchange and web site usage.

Since its implementation, the amount of data in the LAS has doubled. Each month the LAS serves an average of 1.5GB of data to users. This includes all images and text files. Approximately 2000 users access the LAS on a monthly basis. The number of users has increased steadily over the past two years. LAS usage statistics are recorded each month to show how often it is being utilized. These metrics are not the only source of feedback on the LAS. Comments from workshop participants also assist in providing criticism about how to improve the user interface. Improvements will continue for the lifetime of the REASoN project.

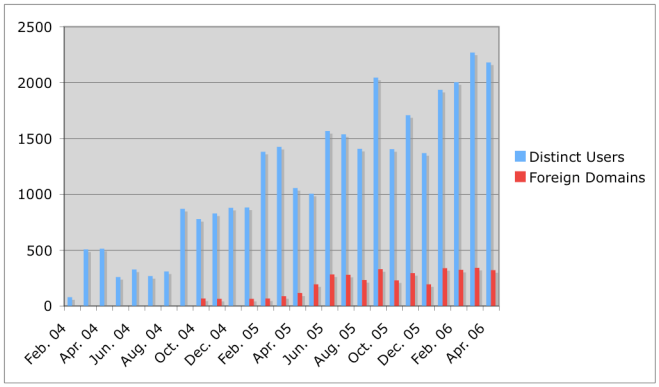


Figure 4. History of Distinct Users and Foreign Domains accessing the MY NASA DATA website since the project's inception.